

An Automatic Tool for Quantification of Nerve Fibres in Corneal Confocal Microscopy Images

Xin Chen, Jim Graham, Mohammad A. Dabbah, Ioannis N. Petropoulos, Mitra Tavakoli, Rayaz A. Malik

Abstract— Objective: We describe and evaluate an automated software tool for nerve fibre detection and quantification in corneal confocal microscopy (CCM) images, combining sensitive nerve-fibre detection with morphological descriptors. **Method:** We have evaluated the tool for quantification of Diabetic Sensorimotor Polyneuropathy (DSPN) using both new and previously published morphological features. The evaluation used 888 images from 176 subjects (84 controls and 92 patients with Type 1 diabetes). The patient group was further subdivided into those with (n=63) and without (n=29) DSPN. **Results:** We achieve improved nerve-fibre detection over previous results (91.7% sensitivity and specificity in identifying nerve-fibre pixels). Automatic quantification of nerve morphology shows a high correlation with previously reported, manually measured, features. ROC analysis of both manual and automatic measurement regimes resulted in similar results in distinguishing patients with DSPN from those without: AUC of about 0.77 and 72% sensitivity-specificity at the equal error rate point. **Conclusion:** Automated quantification of corneal nerves in CCM images provides a sensitive tool for identification of DSPN. Its performance is equivalent to manual quantification, while improving speed and repeatability. **Significance:** Corneal confocal microscopy is a novel in-vivo imaging modality that has the potential to be a non-invasive and objective image biomarker for peripheral neuropathy. Automatic quantification of nerve morphology is a major step forward in the early diagnosis and assessment of progression, and, in particular, for use in clinical trials to establish therapeutic benefit in diabetic and other peripheral neuropathies.

Index Terms— Diabetic Sensorimotor Polyneuropathy, Computer Aided Diagnosis, Corneal Confocal Microscopy, Image Analysis, Nerve Fibre Quantification

I. INTRODUCTION

DIABETIC sensorimotor polyneuropathy (DSPN) is one of most common long term complications of diabetes. Up to

50% of diabetic patients suffer from it [1], and it is estimated that about one in six diabetic patients have chronic painful neuropathy [2]. Several methods are currently used to quantify neuropathy, including clinical scoring of symptoms, quantitative sensory testing, nerve conduction measurements and microscopic measurement of intra-epidermal nerve-fibre density (IENFD) in skin biopsy samples. These methods have their advantages and limitations. Thus, whilst symptoms and signs are directly relevant to the patient and are easily recorded, they are subjective resulting in poor repeatability [3]. Neurophysiology is more objective; however it only assesses large fibres, which constitute a tiny proportion of all the nerve fibres present in a peripheral nerve and has also been shown to have limited reproducibility [4]. The quantification of IENFD in skin biopsies is objective, but is clearly invasive and requires considerable expertise in assessment. There is a need for a rapid, non-invasive assessment that is truly quantitative and assesses small nerve fibres, which are more likely to be involved in neuropathy [5, 6].

Corneal confocal microscopy (CCM) images of nerve fibres are captured from the sub-basal plexus immediately above Bowman's membrane of the cornea by an in-vivo laser confocal microscope. Fig. 1a shows an example image. One of the advantages of CCM is the entirely non-invasive and relatively rapid (about 2 minutes) acquisition of images of small nerve fibres and other corneal structures. Clinical studies [7] have shown that CCM is capable of making quantitative assessment of DSPN and has the potential to be an ideal surrogate endpoint. It has also recently been shown to have a predictive ability in identifying diabetic patients at risk of developing DSPN [8] and has been used in several clinical intervention studies showing nerve-fibre repair [9-11]. Interactive analysis has been used to derive measurements from these images, such as corneal nerve-fibre length (CNFL), corneal nerve-fibre density (CNFD) and corneal nerve branch density (CNBD) [12, 13] (Fig. 1). CNFL is defined as the total length of all nerve fibres visible in the CCM image per square millimetre. CNFD and CNBD are the number of the major nerves (red lines in Fig. 1b) per square millimetre and the number of primary branches emanating from those major nerve trunks (green dots in Fig. 1b) per square millimetre respectively. Although an association has been demonstrated between these quantitative features and the severity of DSPN [7] in cross sectional studies, the manual analysis suffers from the usual problems of being labour-intensive and subjective and therefore raises considerable difficulties, particularly when undertaking longitudinal follow-up studies [14].

- J. Graham is with the Centre for Imaging Sciences, the University of Manchester, UK. E-mail: jim.graham@manchester.ac.uk
 - X. Chen is now at the Division of Imaging Sciences and Biomedical Engineering, Kings College London, UK. E-mail: xin.chen@kcl.ac.uk
 - M. Dabbah is currently at Roke Manor Research Ltd. Romsey, UK
 - I. Petropoulos, M. Tavakoli and Rayaz Malik are with Centre for Endocrinology & Diabetes, Institute of Human Development, Manchester, UK. I. Petropoulos and Rayaz Malik are also with Weill Cornell Medical College in Qatar, Division of Medicine, Doha, Qatar.
- Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

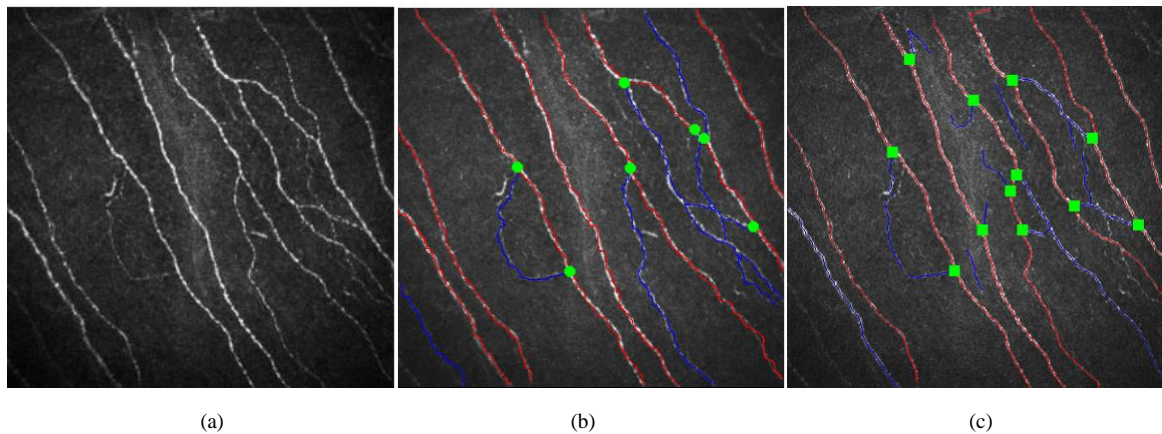


Fig. 1. (a) Original CCM image. (b) Manually quantified CCM image. (c) Automatically quantified CCM image. Red lines represent main nerve fibres, blue lines are branches and green spots indicate branch points on the main nerve trunks. Refer to online coloured version.

Consequently the quantification results show poor reproducibility, especially in CNBD [15]. For the technology to be clinically useful, the analysis of images needs to be done automatically.

Here we describe a fully automatic nerve-fibre detection and quantification system. Fig. 1a indicates that the appearance of nerve fibres in CCM images covers a wide contrast range, with some fibres appearing very faint on a noisy background, whilst other, larger, fibres show strong contrast. A number of studies have presented methods of detecting similar linear structures in different types of images e.g. the detection of blood vessels in retinal images [16], and the detection of curvilinear structure in mammograms [17]. Previous studies aimed at automatic fibre detection in CCM images include Scarpa et al. [18] who described a method for tracing nerve fibres based on automatically initialised seed points, and Holmes et al. [19] who identified fibres based on ridge points. Sindt et al. [20] detected several types of objects visible in CCM images, including dendritic immune cells and wing cells in addition to nerve fibres. Dabbah et al. [21] presented a method of fibre detection based on a multi-scale Gabor filter with responses trained using a neural network. The best detection performances in various applications are achieved using methods based on machine learning, in which features are derived from training images [16, 17, 21].

Following fibre detection, it is required to extract individual fibres, identify branches and quantify appropriate features for classification. A number of studies have investigated the quantification of a variety of image features, describing the morphology of nerve fibres delineated either manually or automatically [13, 19-21]. These studies have shown the relationship between several features, including those listed above, and neuropathic status. None of them, however, has addressed the question of diagnosis of individual subjects.

We have previously described our image filter for enhancing nerve-fibre pixels [21] and reported clinical results of applying this system to DSPN [22]. This paper describes the development of the fibre detection method into a complete tool for measurement of nerve-fibre morphology to act as a diagnostic aid, making three specific contributions over our earlier publications: (1) we compare our fibre detector [21] with another, successful, linear feature descriptor and demonstrate the best reported performance in detecting nerve-

fibre pixels in CCM images. (2) The detailed algorithms for quantification of morphometric features are presented for the first time (only CNFL was used in [21]), including the established features (CNFD, CNFL, CNBD) and new features: Corneal Nerve-Fibre Width Histogram (CNFWH) and Corneal Nerve-Fibre Orientation Histogram (CNFOH). (3) Finally, we report a technical validation of the proposed system based on CCM images obtained from 84 control subjects and 92 type 1 diabetic patients, which, to our knowledge, is the largest dataset in the literature for DSPN diagnosis of individuals.

II. METHODS

A. CCM Images and Manual Measurement

CCM images (Fig. 1(a)) were captured from all participants using the Heidelberg Retina Tomograph Rostock Cornea Module (HRT-III) as described in [13]. The image dimensions are 384×384 pixels with the pixel size of $1.0417 \mu\text{m}$. During the CCM scan, images captured from all corneal layers and six sub-basal images from the right and left eyes were selected for analysis. Criteria for image selection were depth, focus position and contrast. A single experienced examiner, masked from the outcome of the medical and peripheral neuropathy assessment, manually quantified images of all study participants using purpose-written proprietary software (CCMetrics: M. A. Dabbah, Imaging Science, University of Manchester) to delineate main fibres, branch fibres and branch points (red lines, blue lines and green dots respectively in Fig. 1b). The reproducibility and reliability of manual annotation are reported in [15]. The specific parameters measured in each frame were [ENREF 18](#): CNFD, CNFL and CNBD, as described in section I in accordance with our previously published protocol [13].

B. Automated CCM Measurement

The automated CCM measurement process consists of two main steps: nerve-fibre detection and nerve-fibre quantification.

1) Nerve-Fibre Detection:

In this and similar applications [16, 17], methods based on machine learning have been reported to outperform others in detection of curvilinear features. The machine learning method normally consists of two key elements, feature description and classifier training on a set of samples.

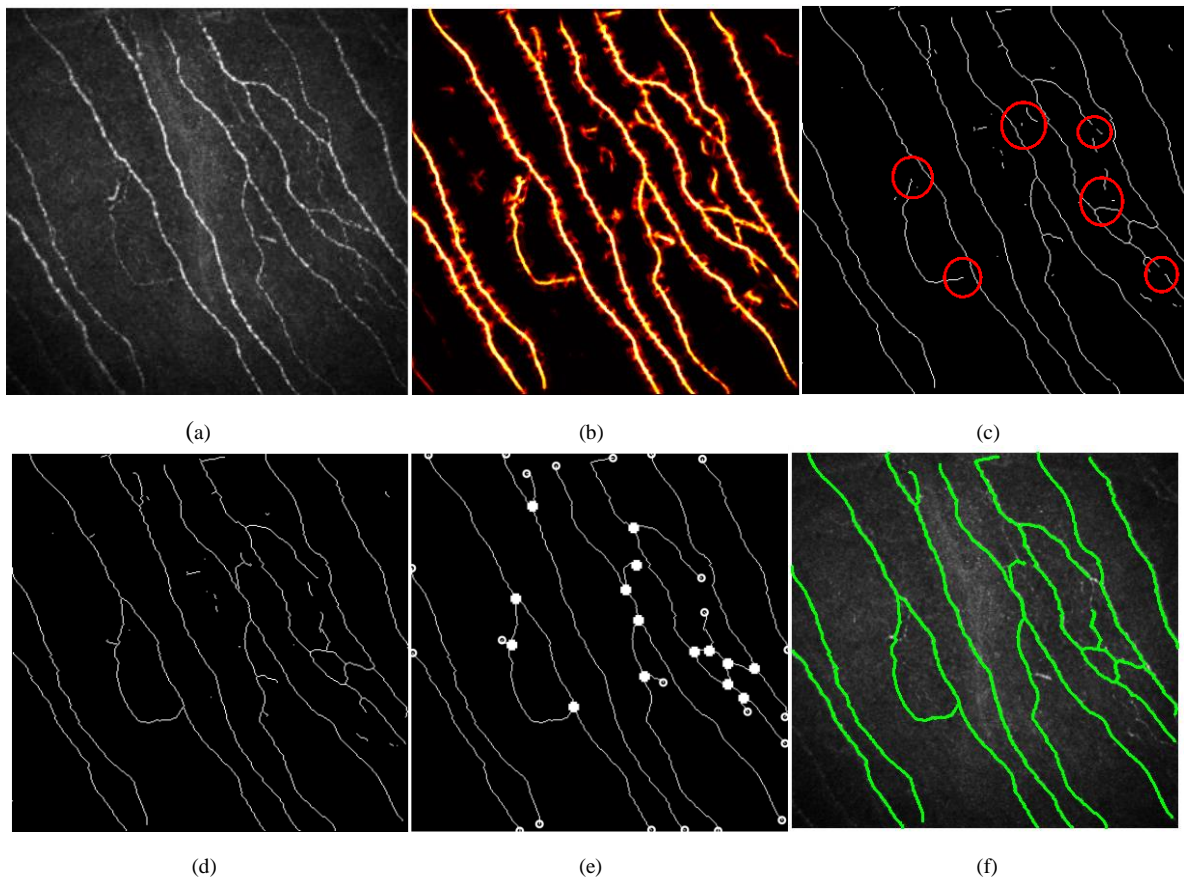


Fig. 2. (a) Original CCM image (b) Response image after nerve-fibre detection and denoising (c) Nerve-fibre skeleton with highlighted weak connection segments (d) Nerve-fibre skeleton after assessment of weak connections. (e) Automatically detected end points (hollow circles) and intersection points (solid circles). (f) Final detected nerve fibres.

For the feature description process, we have implemented and adapted two of the most successful methods [17, 21] for representing curvilinear structures. Dabbah et al. [21] proposed a multi-scale “dual-model filter” (DMF) that combines a foreground model based on a Gabor wavelet with a Gaussian background model that scales the output according to the level of noise. In our implementation, we apply the DMF at eight orientations (suggested in [21]) and at four levels of an image pyramid. Each level is a down-sampled (with smoothing) version of its immediate higher level by a factor of 2. The Gabor wavelet and Gaussian filter covered orientations from 0° to 180° and a range of fibre widths that we found to be sufficient for the CCM images in our study. The DMF method results in 32-dimensional vectors (8 orientations \times 4 scale pyramid levels) to describe features at each pixel location.

Berks et al. [17] described a system that used the dual-tree complex wavelet transform (DWT) [23] for detection of linear structures in mammograms. The DWT combines the outputs of two discrete transforms, using real wavelets differing in phase by 90° , to form the real and imaginary parts of complex coefficients. It provides a directionally selective representation with approximately shift-invariant coefficient magnitudes and local phase information. As in the DMF method, the DWT is applied to a four-level image pyramid. Additionally, the DWT is performed at six different orientations ($\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$, used in [17]) at each pyramid level. The six sub-bands are then

multiplied by $\{i, -i, i, -1, 1, -1\}$ respectively, so that the phase at the centre of the impulse response of each wavelet is zero. Finally, to achieve 180° rotational symmetry, any coefficient with negative imaginary part is replaced with its complex conjugate. Hence from coarse level to fine level of the image pyramid, the DWT results in a 48-element feature vector (4 level image pyramid \times 6 orientation \times 2 magnitude and phase) for each selected pixel location. Both of these detectors outperformed competitors in their respective domains.

In this study we have implemented both detectors in the form proposed by the original authors (number of pyramid levels and orientations) as these produced feature vectors of similar dimension. We then subjected them to a comparative analysis in detecting nerve fibres.

For classifier training, the feature descriptors and their corresponding fibre/non-fibre labels from a set of training samples were used as the inputs to a classifier, which took the form of either a neural network or random forest [24]. The trained classification model was then used for classifying fibre/non-fibre pixels in unseen CCM images.

CCM images have a fairly high level of background noise (see Fig. 2a), which at a fine scale have similar contrast to nerve fibres at random orientations. These may be detected by the trained detectors and are removed by a further denoising step, which iteratively diminishes pixels that are not consistent with the dominant direction over a localised region. The output response image after denoising is shown in Fig. 2b. The

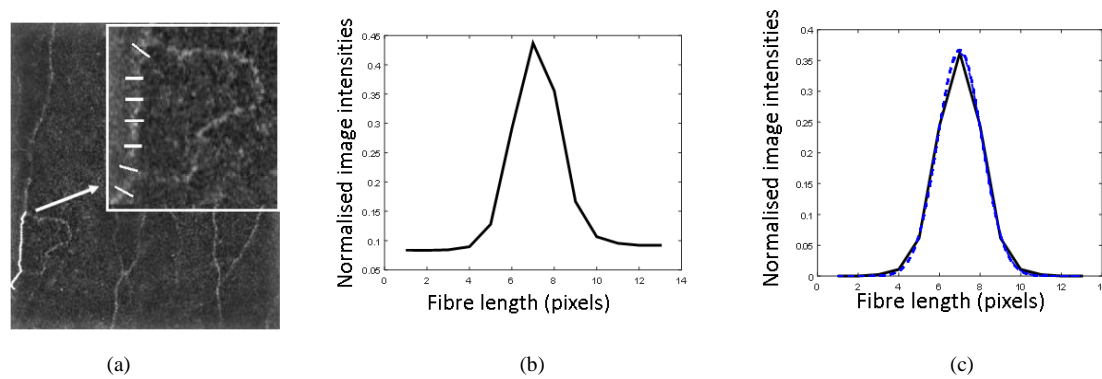


Fig. 3. (a) Original CCM image with a highlighted segment, a selection of orthogonal profile lines are indicated on the enlarged inset. Profiles are calculated at each pixel along the segment. (b) Average of all the profile lines along the whole fibre segment. (c) The symmetric profile of (b) is firstly calculated, and then normalised (Solid line). A Gaussian distribution is fitted for nerve-fibre width estimation (broken line). The final width equals 2.5 times the RMS width (σ) of the fitted Gaussian curve.

evaluation and comparison of different combinations of the two feature descriptors and the two classifiers, before and after denoising, are presented in section III.

Based on the denoised image, a threshold is then applied to generate a binary image. The optimum threshold value is determined by the training and validation experiments described in section III-A. The binary image is then filtered by morphological operators to fill small gaps within nerve fibres and link adjacent structures. The binary structures are thinned to obtain a one-pixel wide skeleton (Fig. 2c). Branch and end points, identified by counting the neighbours of each skeleton point, are each assigned a unique label. For some regions, the evidence for nerve fibres is too weak (as highlighted in Fig. 2c) to be detected by a global threshold. However, the undetected pixels may be important in determining the nerve-fibre connectivity. Hence, for each end point, we extrude 30 pixels along the fibre orientations. The orientation of nerve fibres at each pixel location can be estimated using the second eigenvalue of the Hessian matrix of the response image. If an intersection with another fibre is detected and the average probability from the response image of the extruded pixels is sufficiently high (> 0.2), the extruded line is retained, otherwise it is eliminated (Fig. 2d). Subsequently, independent small segments and short branches that are less than 15 pixels long are removed, and the intersection points (solid circles) and end points (hollow circles) are calculated again as shown in Fig. 2e. The final binary skeleton, as shown in Fig. 2f, is used for total nerve-fibre quantification, described in the next section.

2) Nerve-Fibre Quantification

Fig. 2f shows that the output of fibre detection consists of several networks of interconnected line segments. In order to produce similar results to the manual CNFD, CNFL and CNBD, it is important to identify the main fibres within these networks and the branch points along the main fibres. To connect the appropriate fibre segments together, we generate four $N \times N$ matrices (MI, ML, MW and MO) to store the fibre intensity, fibre length, fibre width (described later in this section) and fibre orientation information respectively for each segment. N is the total number of branch and end points. If the i^{th} and j^{th} end/branch points are connected by a segment, the intensity, width, length and orientation information will be

saved at the $[i, j]$ location of the corresponding matrices; if they are not connected, these elements are zero. The matrices of intensity (MI), length (ML) and width (MW) are symmetric, as the elements at $[i, j]$ and $[j, i]$ should be identical. The $[i, j]$ and $[j, i]$ elements in the orientation matrix MO represent the respective orientations of the opposite ends of the fibre segment.

Identification of the main nerve fibres starts with the most prominent segments: those with greatest length and width. These are identified by multiplying the corresponding elements of MW and ML to produce a new matrix MA. The segments are considered in sequence according to the corresponding values of MA in descending order. There are normally two candidate segments that intersect with the current segment at a branch point. The candidate segments are ranked for the length, orientation difference, intensity and width parameters respectively. The candidate with the highest summed rank is chosen to connect with the current segment. The process continues till an end point is reached. The relevant entry in MA is set to zero and the process continues until no non-zero elements remain in MA. Finally, a list of connected fibres is obtained. Only the fibres with length greater than a threshold are kept as the main fibres. Fig. 1b and 1c respectively show the manual and automatic quantification results of the CCM image in Fig. 1a. The red lines show the principal nerve fibres, which are counted to produce CNFD. The blue lines indicate the secondary nerve fibres, which together with the principal fibres make up CNFL. The green dots are the branch points from the main fibres that are used for CNBD calculation.

Besides the CNFD, CNFL and CNBD features that are readily measured in the manual analysis, automatic quantification is able to calculate additional features. These additional CCM features include the total corneal nerve-fibre area per mm^2 (CNFA), the corneal nerve-fibre width histogram (CNFWH) and the corneal nerve-fibre orientation histogram (CNFOH). These can be calculated if the width and orientation at each nerve-fibre location is known. The orientation is calculated by the Hessian method referred to in section II-B-1. The nerve-fibre width estimation for a particular segment is illustrated in Fig. 3. Fig. 3a shows a highlighted example nerve-fibre segment along with a

magnified version. At each nerve-fibre location, an intensity profile line of length 13 pixels (larger than the thickest fibre) is extracted perpendicular to the nerve-fibre orientation, as indicated by the short straight lines in Fig. 3a. The profiles corresponding to a fibre segment are averaged along the length of the segment to generate a representative profile for the segment, which is then further averaged (Fig. 3b) with its symmetrically inverted profile, smoothed by a three pixel length average filter and normalised (Fig. 3c). Finally a Gaussian curve is fitted to the normalised profile curve (Fig. 3c). The final width of that segment is calculated as 2.5 (empirically determined) times the RMS width of the fitted Gaussian curve. CNFWH is the number of occurrences of different fibre widths in the range between 1 to 8 pixels, at 0.2 pixels interval (36 bins). The CNFA is calculated as sum of fibre width \times fibre length of all the fibre segments in mm^2 . The CNFOH is the number of occurrences of different fibre orientations in the range between 0° to 179° , at 5 degree interval (36 bins).

III. MATERIALS AND EVALUATION

We performed the model training and testing processes on two independent datasets. Dataset 1 contains 200 CCM images which were randomly selected from healthy volunteers and subjects who were diagnosed with type 1 diabetes. This dataset was used for model training and validation for parameter optimisations. The testing stage was conducted on an independent dataset 2 that contained 888 images captured from 176 subjects (84 controls and 92 diabetic patients). The subjects were divided into 3 groups: control ($n=84$), type 1 diabetic patient with no neuropathy ($n=63$) and diabetic patients with neuropathy ($n=29$). The Toronto Diabetic Neuropathy Expert Group (TC) [6] recommendation was followed to define an individual to have DSPN if he/she met both of the following criteria: (1) Abnormal nerve conduction – A peroneal motor nerve conduction velocity of <42 m/s; (2) a symptom or sign of neuropathy, defined as ONE of the following: (a) diabetic neuropathy symptom (DNS) [25] of 1 or more out of 4, (b) neuropathy disability score (NDS) [26] of 3 or more out of 10. These features, along with a number of other clinical and physiological parameters, were measured for each subject [22].

Following the description in section II-A, all images from both dataset 1 and dataset 2 were acquired by the same procedure. They were all manually segmented by a trained clinician (INP). CNFD, CNFL and CNBD were also measured manually in each of the images using the CCMetrics annotation tool (denoted as MCNFD, MCNFL, and MCNBD, respectively).

A. Evaluation of Nerve-Fibre Detection

For the evaluation of nerve-fibre detection, we firstly trained and validated the models based on dataset 1 using a two-fold cross validation. The 200 images in dataset 1 were randomly divided into two groups with 100 images each. Each set served for parameter setting and training based on the other half as a test set. The roles were then reversed. We performed this two-

fold cross validation on the four combinations of feature descriptors (DMF and DWT) and model classifiers [24] (random forest (RFC) and multi-layer perceptron neural network classifiers (NNC)), denoted as DMRF (DMF + RFC), DMNN (DMF + NNC), DTRF (DWT + RFC) and DTNN (DWT + NNC). For each of the combinations, we repeated the two-fold cross validation to investigate the optimum parameter settings by varying the number of training pixels (500, 1000, 2000 pixels randomly selected from each of the foreground and background regions for each image), the number of trees (100, 200 and 500 trees) in RFC and the number of hidden neurons (20, 50 and 100) in the three-layer NNC.

For performance evaluation, as in [21], the response images (before denoising) were thresholded and thinned to one-pixel wide lines. These lines were then compared pixel by pixel to the manually generated skeletons acting as ground-truth, a true positive (TP) being scored if the detected pixel is within a three-pixel tolerance of ground truth and a false positive (FP) if it is outside this tolerance. True negative (TN) and false negative (FN) pixels are recorded if the pixel in the detected image is zero while the ground-truth is zero and one respectively. The three pixel tolerance deals with the imprecision in placing hand-drawn centrelines. By varying the threshold of the response images, ROC curves can be generated for each of the parameter settings. Optimum performance, in terms of specificity and sensitivity at the equal error rate point and computational time, was achieved by using 1000 foreground and background pixels from each image for training, and 200 trees for RFC and 50 hidden neurons for NNC.

In the testing stage, we applied the two optimised models (one from each of the two-fold cross validation runs on dataset 1) to the independent dataset 2. ROC performances were compared between models trained using DMNN, DMRF, DTNN and DTRF. The values of sensitivity and specificity at the equal error rate point for the two models were as follows. DMNN: 0.917 and 0.913, DMRF: 0.912 and 0.908, DTNN:

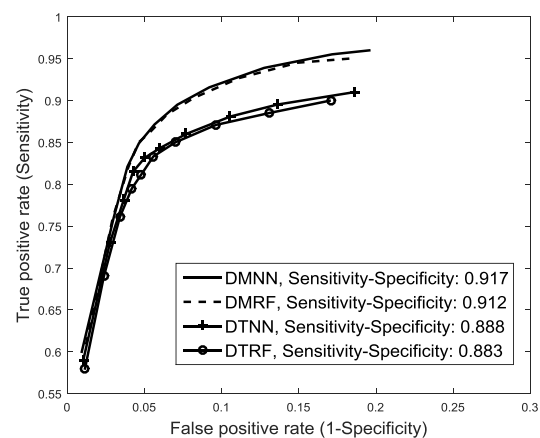


Fig. 4. ROC curves for nerve-fibre detection on dataset 2, using DMNN (Dual Model, Neural Network), DMRF (Dual Model, Random Forest), DTNN (Dual-Tree Wavelet, neural Network) and DTRF (Dual-Tree Wavelet, Random Forest) respectively.

0.888 and 0.882, DTRF: 0.883 and 0.878. In Fig. 4, we show the ROC plots of the model that with the higher performance. These were obtained using the raw detections before denoising to obtain an insight into the underlying detector performance. From the ROC curves, it is clear that the combination based on the DMF outperforms the DTW feature descriptor, having higher sensitivity at any value of specificity.

Although the specificity is a fair measurement for the detection of both background and foreground pixels, the value is dominated by the very high TN count. The absolute value of specificity is potentially misleading, as we have noted in section II-B-1 that the initial detection results in detection of a high number of background pixels that are removed by a subsequent denoising step. We therefore also calculated the False Discovery Rate ($FDR = FP / (FP + TP)$) and the False Negative Rate ($FNR = FN / (FN + TP)$). The smallest (best) FDR/FNR measures for the four methods before and after image denoising are listed in Table I. Since the two cross validation models produce very similar classifications, we only report the results from one of the cross validation models. All four detector/classifier combinations have similar FDR values before denoising and, significantly reduced, after denoising, consistent with the similar specificity values at most values of sensitivity in Fig. 4. The FNR values increase only slightly by denoising, the DM detector achieving better FNR figures. Following denoising there is no real difference between the RF and NN classifiers.

TABLE I
FDR AND FNR FOR THE FOUR COMPARED DETECTORS
BEFORE AND AFTER IMAGE DENOISING.

Method	FDR/FNR Before denoising	FDR/FNR After denoising
DMNN	0.4810/0.2700	0.2013/0.2934
DMRF	0.4828/0.2877	0.2014/0.2983
DTNN	0.4890/0.2961	0.2012/0.3141
DTRF	0.4881/0.3221	0.2013/0.3261

B. Evaluation of Nerve-Fibre Quantification

As observed in section III-A the two cross-validation models produced very similar performance on the independent dataset 2. We chose the detector model with the slightly higher performance as the basis for automated measurements of nerve-fibre parameters, denoted ACNFD, ACNFL and ANCBD. Additionally, total nerve-fibre area, orientation histogram and width histogram were calculated (CNFA, CNFOH and CHWH). For the multidimensional features CNFOH and CNFWH, we investigated the use of the maximum, standard deviation, skewness, kurtosis and logistic regression combining all elements of the histogram feature vectors to represent the feature. The standard deviation of the histogram proved to be the most effective; these are denoted as ASDOH and ASDWH.

For each of the subjects, the average feature values obtained from their CCM images were used. Fig. 5 and Fig. 6 show the box plots of each of the manual and automated CCM features respectively. In these figures, the central red lines are the

median, the edges of the box are the 25th and 75th percentiles ($q1$ and $q3$), and the whiskers extend to the most extreme data points that are not identified as being outliers (within the range $q1 - 1.5(q3 - q1)$ to $q3 + 1.5(q3 - q1)$). The outliers are plotted individually as red dots. A common decreasing trend from control group to neuropathy group can be observed on all manual and automated CCM features. The values of the manually generated measurements are higher than those generated automatically. One reason for this is that the manual tracing process deviates from the exact fibre path (Fig 1(b)), resulting in a larger CNFL value. Additionally, the automated method is less effective than human annotators at connecting weak branches, resulting in generally higher CNFD and CNBD values for the manual analysis. However, the important point is the relative correlation between manual and automated measures across the control and patient groups. The Pearson correlation coefficients between automatically and manually derived CNFL, CNFD and CNBD measurements were 0.861, 0.859 and 0.701 respectively. The lower correlation in the case of CNBD measurement is due to poor reproducibility in the manual measurement of this feature. This has been reported in [15] and arises from the subjective judgement required for identifying branch points.

We used both the ANOVA test [27] and ROC analysis to demonstrate the capability of using the CCM image features to discriminate between control and non-neuropathic groups, and between non-neuropathic and neuropathic patients, as defined by the Toronto Criteria.

Tables II and III show the respective ANOVA p-values, the area under the ROC curve (AUC) measures and sensitivity and specificity values calculated at the equal error point (EEP) of the ROC curves. We also experimented with different combinations of features, from both manual and automated analysis, using logistic regression in a leave-one-out manner. In these experiments each subject was predicted by the logistic regression model built from the remaining $n-1$ subjects, where n is the total number of subjects in both groups. ROC measures for the combinations of all manual features or all automated features are listed in Table II and Table III along with the single-feature measures. The confidence intervals for the combined methods indicate that the combination results in a discriminating power indistinguishable from the best manual or automatic methods respectively. This would indicate that the features are accessing the same underlying information about each of the groups. It is unsurprising that there should be dependency between total fibre length and fibre density or fibre area.

IV. DISCUSSION

A number of studies have shown the features extracted from Corneal Confocal Microscopy images are associated with the severity of diabetic peripheral neuropathy [7, 12, 13] and the potential of CCM to quantify severity of neuropathy and assess therapeutic benefit has been demonstrated [28]. In this paper, we have described the details of a complete system for measurement of CCM images to enable discrimination between control and diabetic subjects and between diabetic subjects with and without neuropathy.

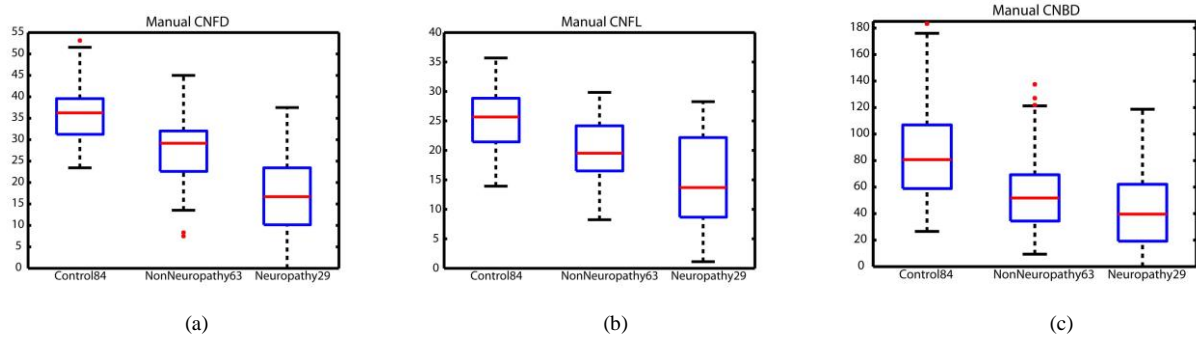


Fig. 5. Boxplots of manually measured features for control, non-neuropathy and neuropathy groups (a) MCNFD (b) MCNFL (c) MCNBD.

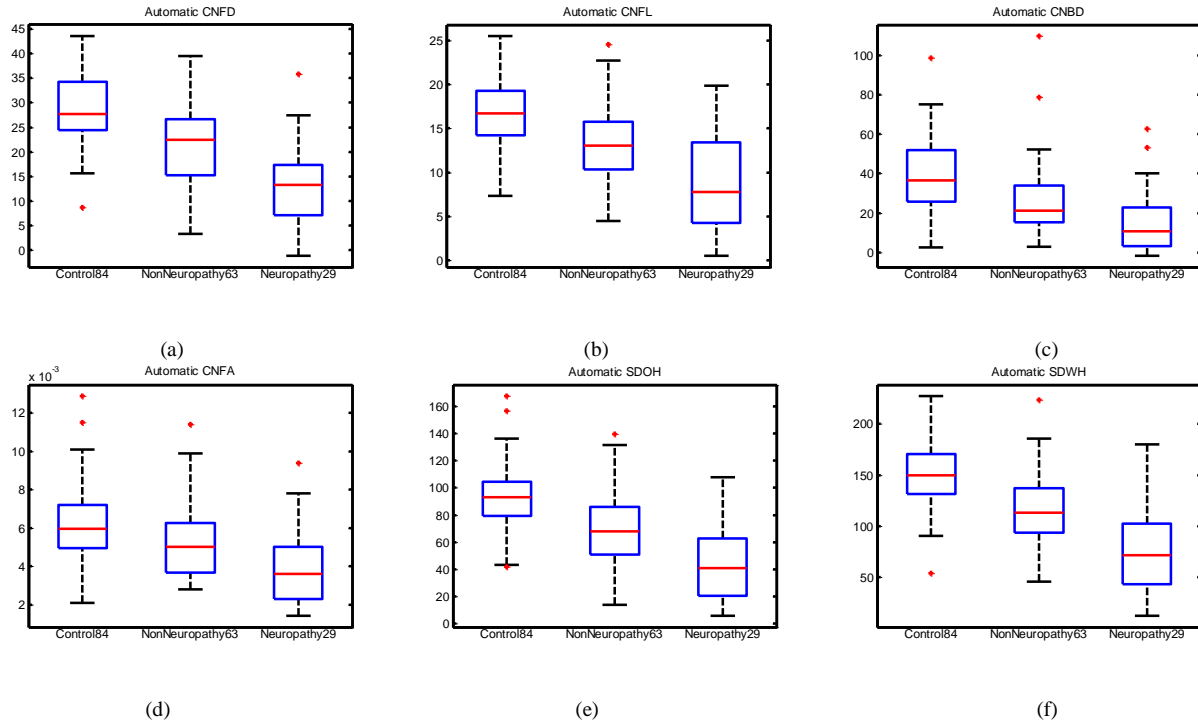


Fig. 6. Boxplots of automatically measured features for control, non-neuropathy and neuropathy groups in dataset 2 (a) ACNFD (b) ACNFL (c) ACNBD (d) ACNFA (e) ASDOH (f) ASDWH.

Petropoulos et al. [22] reported a clinical evaluation study comparing the system described in this paper with manual analysis of CCM images and a broader range of subjective and objective clinical assessment methods, including the Neuropathy Symptom Profile, vibration perception thresholds, cool and warm thermal thresholds, and cold and heat induced pain. CCM features, measured both automatically and manually, were found to be significantly correlated with these methods. They noted that the automatic analysis of CCM images was significantly faster than manual analysis, taking 10-22s per image, depending on the density of fibres, as opposed to 2-7 minutes.

Based on the well-established Toronto Criteria, we show that both manual and automated CCM features discriminate diabetic patients with and without neuropathy. Manual and automatic measurement regimes result in broadly similar results: about 0.77 AUC value and 73% sensitivity-specificity at the equal error rate point. There were no significant differences between the ROCs of manual (MCNFD) and automated measurements (e.g. $p=0.44$ and 0.55 for ACNFD

and SDWH respectively).

Corneal confocal microscopy has shown considerable success in translation to the assessment of other neuropathies including Fabry disease [29], ISFN [30], CMT1A [31], sarcoidosis [32]. Automated quantification of corneal nerves provides a major step forward in the early diagnosis and assessment of progression, but in particular for use in clinical trials to establish therapeutic benefit in diabetic and other peripheral neuropathies.

The automatic quantification software can be requested freely from [33] for research purposes. It is currently being used by over 40 research groups worldwide to investigate potential relationships between CCM features and different types of neuropathy [34].

V. CONCLUSION

We have presented a technical evaluation of a complete system that is able to automatically quantify six different types of nerve-fibre features in CCM images. We have proposed an optimum configuration for detection of nerve fibres based on a

TABLE II

AUC, 95% CONFIDENCE INTERVAL VALUES AND SENSITIVITY-SPECIFICITY AT THE EQUAL-ERROR POINT (EEP) FOR MANUAL AND AUTOMATED CCM FEATURES FOR DISCRIMINATION BETWEEN CONTROL SUBJECTS AND DIABETIC PATIENTS WITHOUT DSPN.

CCM FEATURES	AUC	95% CI	SENSITIVITY SPECIFICITY AT EEP	P-VALUE OF ANOVA
MCNFD	0.8063	[0.73 0.88]	0.7460	<0.0001
MCNFL	0.7627	[0.68 0.84]	0.6825	<0.0001
MCNBD	0.7492	[0.67 0.83]	0.6984	<0.0001
ACNFD	0.7401	[0.66 0.82]	0.7305	<0.0001
ACNFL	0.7766	[0.70 0.85]	0.7613	<0.0001
ACNBD	0.7103	[0.63 0.80]	0.6414	<0.0001
CNFA	0.6837	[0.59 0.77]	0.6601	0.0002
SDOH	0.7671	[0.69 0.85]	0.7002	<0.0001
SDWH	0.7798	[0.71 0.86]	0.7402	<0.0001
COMBINED MANUAL	0.7940	[0.72 0.87]	0.7143	-
COMBINED AUTOMATED	0.7373	[0.67 0.83]	0.7009	-

TABLE III

AUC, 95% CONFIDENCE INTERVAL VALUES AND SENSITIVITY-SPECIFICITY AT THE EQUAL-ERROR POINT (EEP) FOR MANUAL AND AUTOMATED CCM FEATURES FOR DISCRIMINATION BETWEEN NON-NEUROPATHIC AND NEUROPATHIC GROUPS OF DIABETIC PATIENTS.

CCM FEATURES	AUC	95% CI	SENSITIVITY SPECIFICITY AT EEP	P-VALUE OF ANOVA
MCNFD	0.7890	[0.68 0.90]	0.7241	<0.0001
MCNFL	0.7137	[0.59 0.83]	0.6552	0.001
MCNBD	0.6136	[0.49 0.74]	0.5862	0.081
ACNFD	0.7600	[0.65 0.87]	0.6482	<0.0001
ACNFL	0.7576	[0.65 0.88]	0.6186	<0.0001
ACNBD	0.6801	[0.56 0.80]	0.5798	0.002
CNFA	0.7601	[0.64 0.87]	0.7301	<0.0001
SDOH	0.7799	[0.68 0.90]	0.6907	<0.0001
SDWH	0.7709	[0.67 0.88]	0.7219	<0.0001
COMBINED MANUAL	0.7843	[0.68 0.89]	0.7100	-
COMBINED AUTOMATED	0.7419	[0.63 0.86]	0.6779	-

previously reported foreground and background model trained with a neural network. The automatic quantification results show a high correlation with manually measured CCM features (CNFL, CNFD and CNBD). The results also show significant differences (p-values of ANOVA test in table II) between the control and non-neuropathic group, indicating the system's ability to detect early signs of change from a healthy to a diabetic condition. The automated system is able to produce additional CCM features that measure the area, width and orientation of the nerve fibres (CNFA, CNFWH and CNFOH). All these new measures show significant differences between the non-neuropathic and neuropathic groups (p-values of ANOVA test in table III), with some features achieving 72% sensitivity-specificity at the equal error rate point, indicating the capacity to identify individuals suffering from neuropathy. The advantages in time labour and reproducibility suggest that automatically measured features may be used as a new, non-invasive method for diagnosing diabetic peripheral neuropathy, providing information on small nerve-fibre damage that is not accessible by most currently used methods. The only method in current clinical use that addresses small fibre damage is the intra-epidermal nerve-fibre density (IENFD) measure, which is invasive, requiring a skin biopsy, and currently cannot be evaluated automatically. We have recently shown [35] that analysis of CCM features has favourable diagnostic efficacy to IENFD (AUC of 0.66)

ACKNOWLEDGMENT

This research was funded by awards from: National Institutes of Health (R105991) and Juvenile Diabetes Research Foundation International (27-2008-362).

REFERENCES

- [1] A. J. Boulton, "Management of Diabetic Peripheral Neuropathy," *Clinical Diabetes*, vol. 23, no. 1, pp. 9-15, 2005.
- [2] C. Daousi, I. A. MacFarlane, A. Woodward, T. J. Nurmikko, P. E. Bundred, and S. J. Benbow, "Chronic painful peripheral neuropathy in an urban community: a controlled comparison of people with and without diabetes," *Diabetic Medicine*, vol. 21, no. 9, pp. 976-982, 2004.
- [3] P. J. Dyck, C. J. Overland, P. A. Low, W. J. Litchy, J. L. Davies, P. C. O'Brien, C. v. N. T. Investigators, J. W. Albers, H. Andersen, C. F. Bolton, J. D. England, C. J. Klein, J. G. Llewellyn, M. L. Mauermann, J. W. Russell, W. Singer, A. G. Smith, S. Tesfaye, and A. Vella, "Signs and symptoms versus nerve conduction studies to diagnose diabetic sensorimotor polyneuropathy: CI vs. NPhys trial," *Muscle Nerve*, vol. 42, no. 2, pp. 157-164, 2010.
- [4] P. J. Dyck, J. W. Albers, J. Wolfe, C. F. Bolton, N. Walsh, C. J. Klein, A. J. Zaffit, J. W. Russell, K. Thomas, J. L. Davies, R. E. Carter, L. J. Melton, W. J. Litchy, and C. v. N. T. Investigators, "A trial of proficiency of nerve conduction: greater standardization still needed," *Muscle nerve*, vol. 48, no. 3, pp. 369-374, 2013.
- [5] P. J. Dyck, J. E. Norell, H. Tritschler, K. Schuette, R. Samigullin, D. Ziegler, E. J. Bastyr, W. J. Litchy, and P. C. O'Brien, "Challenges in Design of Multicenter Trials: Endpoints Assessed Longitudinally for Change and Monotonicity," *Diabetes Care*, vol. 30, pp. 2619-2625, 2007.
- [6] S. Tesfaye, A. J. Boulton, P. J. Dyck, R. Freeman, M. Horowitz, P. Kempler, G. Lauria, R. A. Malik, V. Spallone, A. Vinik, L.

- Bernardi, and P. Valensi, "on behalf of the Toronto Diabetic Neuropathy Expert Group, Diabetic neuropathies: Update on definitions, diagnostic criteria, estimation of severity, and treatments," *Diabetes Care*, vol. 33, no. 10, pp. 2285-2293, 2010.
- [7] P. Hossain, A. Sachdev, and R. A. Malik, "Early detection of diabetic peripheral neuropathy with corneal confocal microscopy," *The Lancet*, vol. 366, no. 94, pp. 1340-1343, 2005.
- [8] N. Pritchard, K. Edwards, A. W. Russell, B. A. Perkins, R. A. Malik, and N. Efron, "Corneal confocal microscopy predicts 4-year incident peripheral neuropathy in type 1 diabetes," *Diabetes Care*, vol. 38, no. 4, pp. 671-675, 2015.
- [9] M. Brines, A. N. Dunne, M. V. Velzen, P. L. Proto, C. G. Ostenson, R. I. Kirk, I. Petropoulos, S. Javed, R. A. Malik, A. Cerami, and A. Dahan, "ARA 290, a non-erythropoietic peptide engineered from erythropoietin, improves metabolic control and neuropathic symptoms in patients with type 2 diabetes," *Molecular Medicine*, vol. 6, 2014.
- [10] M. Tavakoli, M. Mitu-Pretorian, I. N. Petropoulos, H. Fadavi, O. Asghar, U. Alam, G. Ponirakis, M. Jeziorska, A. Marshall, N. Efron, A. J. Boulton, T. Augustine, and R. A. Malik, "Corneal confocal microscopy detects early nerve regeneration in diabetic neuropathy after simultaneous pancreas and kidney transplantation," *Diabetes Care*, vol. 62, no. 1, pp. 254-260, 2013.
- [11] S. Azmi, M. Ferdousi, I. N. Petropoulos, G. Ponirakis, H. Fadavi, M. Tavakoli, U. Alam, and W. Jones, "Corneal confocal microscopy shows an improvement in small-fiber neuropathy in subjects with type 1 diabetes on continuous subcutaneous insulin infusion compared with multiple daily injection," *Diabetes Care*, vol. 38, no. 1, pp. e3-e4, 2015.
- [12] M. Tavakoli, C. Quattrini, C. Abbott, P. Kallinikos, A. Marshall, J. Finnigan, P. Morgan, N. Efron, A. Boulton, and R. Malik, "Corneal Confocal Microscopy: A Novel Non-invasive Test to Diagnose and Stratify the Severity of Human Diabetic Neuropathy," *Diabetes Care*, vol. 33, no. 8, pp. 1792-1797, 2010.
- [13] R. A. Malik, P. Kallinikos, C. A. Abbott, C. H. M. v. Schie, P. Morgan, N. Efron, and A. J. M. Boulton, "Corneal confocal microscopy: a noninvasive surrogate of nerve fibre damage and repair in diabetic patients," *Diabetologia*, vol. 46, no. 5, pp. 683-688, 2003.
- [14] C. Dehghani, N. Pritchard, K. Edwards, D. Vagenas, A. W. Russell, R. A. Malik, and N. Efron, "Morphometric stability of the corneal subbasal nerve plexus in healthy individuals: 1 3-year longitudinal study using corneal confocal microscopy," *Invest Ophthalmol Visual Science*, vol. 55, no. 5, pp. 3195-3199, 2014.
- [15] I. Petropoulos, T. Manzoor, P. Morgan, H. Fadavi, O. Asghar, U. Alam, G. Ponirakis, M. Dabbah, X. Chen, J. Graham, M. Tavakoli, and R. Malik, "Repeatability of In Vivo Corneal Confocal Microscopy to Quantify Corneal Nerve Morphology," *Cornea*, vol. 32, no. 5, pp. 83-89, 2013.
- [16] M. Niemeijer, J. J. Staal, B. v. Ginneken, M. Loog, and M. D. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," *SPIE Medical Imaging*, vol. 5370, pp. 648-656, 2004.
- [17] M. Berks, Z. Chen, S. Astley, and C. Taylor, "Detecting and classifying linear structures in mammograms using random forests," *IPMI 11 proceedings of the 22nd international conference on information processing in medical imaging*, pp. 510-524, 2011.
- [18] F. Scarpa, E. Grisan, and A. Ruggeri, "Automatic Recognition of Corneal Nerve Structures in Images from Confocal Microscopy," *Investigative Ophthalmology and Visual Science*, vol. 49, no. 11, pp. 4801-4807, 2008.
- [19] T. Holmes, M. Pellegrini, C. Miller, T. Epplin-Zapf, S. Larkin, S. Luccarelli, and G. Staurengi, "Automated Software Analysis of Corneal Micrographs for Peripheral Neuropathy," *Investigative Ophthalmology and Visual Science*, vol. 51, no. 9, pp. 4480-4491, 2010.
- [20] C. W. Sindt, B. Lay, H. Bouchard, and J. R. Kern, "Rapid Image Evaluation System for Corneal In Vivo Confocal Microscopy," *Cornea*, vol. 32, no. 4, pp. 460-465, 2013.
- [21] M. A. Dabbah, J. Graham, I. N. Petropoulos, M. Tavakoli, and R. A. Malik, "Automatic analysis of diabetic peripheral neuropathy using multi-scale quantitative morphology of nerve fibres in corneal confocal microscopy imaging," *Medical Image Analysis*, vol. 15, pp. 738-747, 2011.
- [22] I. N. Petropoulos, U. Alam, H. Fadavi, A. Marshal, O. Asghar, M. A. Dabbah, X. Chen, J. Graham, G. Ponikaris, A. J. M. Boulton, M. Tavakoli, and R. A. Malik, "Rapid automated diagnosis of diabetic peripheral neuropathy with in vivo corneal confocal microscopy," *Investigative Optics and Visual Science*, vol. 55, pp. 2071-2078, 2014.
- [23] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234-253, 2001.
- [24] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159-190, 2006.
- [25] J. W. G. Meijer, A. J. Smit, E. V. Sonderen, J. W. Groothoff, W. H. Eisma, and T. P. Links, "Symptom scoring systems to diagnose distal polyneuropathy in diabetes: the Diabetic Neuropathy Symptom score," *Diabetic Medicine*, vol. 19, no. 11, pp. 962-965, 2002.
- [26] M. J. Young, A. J. M. Boulton, A. F. Macleod, D. R. R. Williams, and P. H. Sonksen, "A multicentre study of the prevalence of diabetic peripheral neuropathy in the United Kingdom hospital clinic population," *Diabetologia* vol. 36, pp. 150-154, 1993.
- [27] K. Wallis, "Use of ranks in on-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583-621, 1952.
- [28] T. Holmes, M. Pellegrini, C. Miller, T. Epplin-Zapf, S. Larkin, S. Luccarelli, and G. Staurengi, "Automated Software Analysis of Corneal Micrographs for Peripheral Neuropathy," *Investigative Ophthalmology and Visual Science*, vol. 51, no. 9, pp. 4480-4491, 2010.
- [29] M. Tavakoli, A. Marshall, L. Thompson, M. Kenny, S. Waldek, N. Efron, and R. A. Malik, "Corneal confocal microscopy: a novel noninvasive means to diagnose neuropathy in patients with Fabry disease," *Muscle Nerve*, vol. 40, no. 6, pp. 976-84, Dec, 2009.
- [30] M. Tavakoli, A. Marshall, R. Pitceathly, H. Fadavi, D. Gow, M. E. Roberts, N. Efron, A. J. Boulton, and R. A. Malik, "Corneal confocal microscopy: a novel means to detect nerve fibre damage in idiopathic small fibre neuropathy," *Exp Neurol*, vol. 223, no. 1, pp. 245-50, May, 2010.
- [31] M. Tavakoli, A. Marshall, S. Banka, I. N. Petropoulos, H. Fadavi, H. Kingston, and R. A. Malik, "Corneal confocal microscopy detects small-fiber neuropathy in Charcot-Marie-Tooth disease type 1A patients," *Muscle Nerve*, vol. 46, no. 5, pp. 698-704, Nov, 2012.
- [32] M. van Velzen, L. Heij, M. Niesters, A. Cerami, A. Dunne, A. Dahan, and M. Brines, "ARA 290 for treatment of small fiber neuropathy in sarcoidosis," *Expert Opin Investig Drugs*, vol. 23, no. 4, pp. 541-50, Apr, 2014.
- [33] Accmetrics. "<http://www.medicine.manchester.ac.uk/ena/>."
- [34] ENAGroup. "<http://www.human-development.manchester.ac.uk/ena/ACCMetricsuserinstructions/>."
- [35] X. Chen, J. Graham, M. Dabbah, I. Petropoulos, G. Ponirakis, O. Asghar, U. Alam, A. Marshall, H. Fadavi, M. Ferdousi, S. Azmi, M. Tavakoli, N. Efron, M. Jeziorska, and R. Malik, "Small nerve fiber quantification in the diagnosis of diabetic sensorimotor polyneuropathy: comparing corneal confocal microscopy with intraepidermal nerve fiber density," *Diabetes Care*, vol. 38, no. 6, pp. 1138-1144, 2015.